"Express Mail" Mailing Label No. EL339309723US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s):	Catherine Anne Abbott et al.	)	
Application No.	10/070,464	)	
Filed:	July 18, 2002	)	Group Art Unit: 1652
For:	DIPEPTIDYL PEPTIDASES	)	
Examiner:	Sheridan L. Swope	)	Attorney Docket No. FSCB-100

LETTER

Commissioner for Patents
Washington, D. C. 20231

Sir:

Transmitted under the cover of this Letter is a certified copy of Australian Patent Application No. PQ 2762, filed September 10, 1999, the priority of which is claimed in the above-identified application.

Please charge any fees concerning this matter to our Deposit Account No. 15-0508.

Respectfully submitted,

April 23, 2003

By _____
Talivaldis Cepuritis (Reg. No. 20,818)

OLSON & HIERL, LTD.
20 North Wacker Drive
36th Floor
Chicago, Illinois 60606
(312) 580 1180

Patent Office
Canberra

I, TERESA KOLODZIEJCZYK, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PQ 2762 for a patent by THE UNIVERSITY OF SYDNEY as filed on 10 September 1999.

WITNESS my hand this
Ninth day of April 2003

TERESA KOLODZIEJCZYK
TEAM LEADER EXAMINATION
SUPPORT AND SALES

AUSTRALIA

Patents Act 1990

PROVISIONAL SPECIFICATION

Applicant(s):

THE UNIVERSITY OF SYDNEY

Invention Title:

PEPTIDE AND NUCLEIC ACID MOLECULE

The invention is described in the following statement:

TITLE

Peptide and Nucleic Acid Molecule

FIELD OF INVENTION:

The invention relates to a peptide which is capable of cleaving a peptide bond, to a nucleic acid molecule encoding the peptide, to a vector and a cell comprising the nucleic acid molecule, to a composition or a kit comprising the peptide, to a method of making the peptide, to an antibody which binds the peptide and to a method of cleaving a peptide bond using the peptide.

BACKGROUND OF THE INVENTION

Serine proteases are a family of protein cleaving enzymes. Members of this family have distinct substrate specificity. The prolyl oligopeptidases, dipeptidyl peptidase 4 (DPP4) and fibroblast activation protein (FAP) are serine proteases. DPP4 has substrate specificity for peptides which contain the di-peptide sequence, Ala-Pro, and cleaves a peptide which contains the di-peptide by hydrolysis of a peptide bond which is located C-terminal adjacent to proline in the di-peptide. DPP4 also has substrate specificity for peptides which contain the di-peptide sequence, Gly-Pro, and cleaves a peptide which contains this di-peptide by hydrolysis of a peptide bond which is located C-terminal adjacent to proline in the di-peptide. FAP has a substrate specificity which is similar to the specificity of DPP4, although FAP also has gelatinase activity.

SUMMARY OF THE INVENTION

The inventors have isolated and characterised a new prolyl oligopeptidase and the gene encoding it. The inventors have named the new prolyl oligopeptidase DPP4L1.

As described herein, the substrate specificity of DPP4L1 is distinct from the substrate specificity of other prolyl oligopeptidases.

In one aspect, the invention provides a peptide which is capable of cleaving a peptide bond which is C-terminal adjacent to proline in the sequence Ala-Pro, and which is not capable of cleaving a peptide bond which is C-terminal adjacent to proline in the sequence Gly-Pro.

The capacity of DPP4L1 to cleave, or in other words, hydrolyse a peptide bond which is C-terminal adjacent to proline in the dipeptide sequence Ala-Pro shows that DPP4L1 is a prolyl oligopeptidase. The inability of DPP4L1 to cleave a peptide bond which is C-terminal adjacent to proline in the dipeptide sequence Gly-Pro shows that DPP4L1 is a prolyl oligopeptidase with a substrate specificity which is distinguished from other prolyl oligopeptidases.

The capacity of a prolyl oligopeptidase to cleave a peptide bond which is C-terminal adjacent to proline in the di-peptide sequence Ala-Pro, or Gly-Pro, can be determined by standard techniques as described herein. For example, the capacity to cleave a peptide bond which is C-terminal adjacent to proline in the di-peptide sequence Ala-Pro can be determined by observing hydrolysis of a peptide bond which is C-terminal adjacent to proline in the molecule Ala-Pro-p-nitroanilide. The capacity to cleave a peptide bond which is C-terminal adjacent to proline in the dipeptide sequence Gly-Pro can be determined by observing hydrolysis of the peptide bond which is C-terminal adjacent to proline in the molecule Gly-Pro-p-nitroanilide. In one embodiment, the peptide of the first aspect of the invention is capable of cleaving the peptide bond C-terminal adjacent to proline in the compound Ala-Pro-p-nitroanilide and is not capable of cleaving the peptide

bond C-terminal adjacent to proline in a compound selected from the group of compounds consisting of Gly-Pro-p-nitroanilide, Gly-Arg-p-nitroanilide, Gly-Pro-p-toluene sulphonate and Gly-Pro-7-amino-4-triflouromethyl coumarin.

The inventors believe that an amino acid sequence, Gly-Trp-Ser-Tyr-Gly-Gly which is comprised in the amino acid sequence of DPP4L1 described herein, is likely to be involved in the enzymatic activity of DPP4L1. The inventors further believe that the amino acid sequences, Leu-Asp-Glu-Asn-Val-His-Phe-Ala-His and Glu-Arg-His-Ser-Ile-Arg which are also comprised in the amino acid sequence of DPP4L1 described herein, are likely to be involved in the enzymatic activity of DPP4L1. Thus in another embodiment, the peptide of the first aspect of the invention comprises an amino acid sequence Gly-Trp-Ser-Tyr-Gly-Gly. In another embodiment, the peptide comprises an amino acid sequence Leu-Asp-Glu-Asn-Val-His-Phe-Ala-His. In another embodiment, the peptide comprises an amino acid sequence Glu-Arg-His-Ser-Ile-Arg.

The biochemical characterisation of DPP4L1 described herein shows that DPP4L1 consists of 882 amino acids and has a molecular weight of about 100kDa. Thus in another embodiment, the peptide of the first aspect of the invention consists of about 882 amino acids and has a molecular weight of about 100kDa.

The inventors recognise that by using standard techniques it is possible to generate a peptide which is a truncated form of DPP4L1 and which retains the substrate specificity of DPP4L1. Thus it is recognised that a peptide which has the substrate specificity of DPP4L1 may consist of less than 882 amino acids, or may have a molecular weight of less than 100kDa.

As described herein, the amino acid sequence of DPP4L1 which is predicted from the nucleotide sequence of the nucleic acid molecule which encodes DPP4L1 does not contain a consensus sequence for N-linked glycosylation. Therefore the inventors believe that it is unlikely that DPP4L1 is associated with N-linked glycosylation. In this regard, DPP4L1 is distinguished from other prolyl oligopeptidases which contain between 6 and 9 consensus sequences for N-linked glycoyslation. Thus in a further embodiment, an asparagine residue in the amino acid sequence of the peptide of the first aspect of the invention is not linked to a carbohydrate molecule.

The analysis of DPP4L1 expression described herein shows that it is likely that DPP4L1 is expressed as a cytoplasmic protein. The expression of DPP4L1 is therefore distinguished from other prolyl oligopeptidases, which are expressed on the cytoplasmic membrane, or in other words, the cell surface membrane. Thus in another embodiment, the peptide of the first aspect of the invention is not expressed on a cell surface membrane of a cell.

The inventors believe that a peptide which has the substrate specificity of DPP4L1 can be generated which has the amino acid sequence of DPP4L1 described herein and which contains one or more amino acid deletions, substitutions or insertions of that amino acid sequence. It is expected that a peptide which is at least 51% homologous to the amino acid sequence of DPP4L1 described herein, or which is at least 27% identical to the amino acid sequence of DPP4L1, will retain the substrate specificity of DPP4L1. The % homology can be determined by use of the program/algorithm "GAP" which is available from Genetics Computer Group(GCG),Wisconsin. Thus in another embodiment, the peptide of the first aspect of the

invention has an amino acid sequence which is at least 50% homologous to the amino acid sequence of DPP4L1.

As described herein the inventors characterised the nucleotide sequence of the nucleic acid molecule encoding DPP4L1 and from this, were able to predict the amino acid sequence of DPP4L1. The amino acid sequence of DPP4L1 is shown in Figure 1. In an embodiment, the peptide of the first aspect of the invention has the amino acid sequence shown in Figure 1.

The inventors recognise that DPP4L1 may be fused, or in other words, linked to a further amino acid sequence to form a fusion protein which retains substrate specificity of DPP4L1. An example of a fusion protein is described herein which comprises the amino acid sequence of DPP4L1 which is linked to a further "tag" sequence which consists of an amino acid sequence encoding the V5 epitope and a His tag. An example of another fusion protein which comprises the amino acid sequence of DPP4L1 is a GST fusion protein. Thus in another embodiment, the peptide of the first aspect of the invention is linked to a further amino acid sequence.

The inventors further recognise that the amino acid sequence of DPP4L1 shown in Figure 1 may be comprised in a polypeptide so that the polypeptide has the substrate specificity of DPP4L1. The polypeptide may be useful, for example, to alter the protease susceptibility of the DPP4L1 amino acid sequence. Thus in another embodiment, the peptide of the first aspect of the invention is comprised in a polypeptide which has the substrate specificity of DPP4L1.

In a second aspect, the invention relates to a nucleic acid molecule which encodes a peptide according to the first aspect of the invention.

As described herein, the inventors believe that the gene which encodes DPP4L1 is located at band q22 on human chromosome 15. The location of the DPP4L1 gene is distinguished from genes encoding other prolyl oligopeptidases, which are located on chromosome 2, at bands 2q24.3 and 2q23, or chromosome 7. Thus in an embodiment, the nucleic acid molecule of the second aspect of the invention is capable of hybridising to a gene which is located at band q22 on human chromosome 15.

The inventors have characterised the nucleotide sequence of the nucleic acid molecule encoding DPP4L1. The nucleotide sequence of the nucleic acid molecule encoding DPP4L1 is shown in Figure 1. Thus in an embodiment, the nucleic acid molecule of the second aspect of the invention has the nucleotide sequence shown in Figure 1.

The inventors recognise that a nucleic acid molecule which has the nucleotide sequence shown in Figure 1 could be made by producing only the fragment of the nucleotide sequence which is translated. Thus in an embodiment, the nucleic acid molecule of the second aspect of the invention does not contain 5' or 3' untranslated nucleotide sequences.

As described herein, the inventors observed at least three splice variants of DPP4L1 RNA which are of from 2.6 to 3.1 kb in length. As a frame shift mutation or termination signal was not observed in the nucleotide sequence of these splice variants, and as the coding sequence of two of the splice variants include sequence which encodes the DPP4L1 amino acid sequence which is believed to be associated with enzymatic activity, the inventors believe that the splice

variants are likely to have the substrate specificity of DPP4L1. Thus in another embodiment, the nucleic acid molecule of the second aspect of the invention is a fragment of the nucleotide sequence of DPP4L1 shown in Figure 1 which is about 2.6 to 3.1 kb in length and which encodes a peptide according to the first aspect of the invention.

In another embodiment, the nucleic acid molecule of the second aspect of the invention is selected from the group of nucleic acid molecules consisting of T21, T8, Race product, ATCd3-2-1 and ATCd3-3-10, as shown in Figure 1.

In a third aspect the invention provides a vector which comprises a nucleic acid molecule according to the second aspect of the invention.

In one embodiment, the vector of the third aspect of the invention is capable of replication in a COS-7 cell or E.coli. In another embodiment, the vector is selected from the group consisting of λTripleEx, pTripleEx, pGEM-TEasyRVector and pCDNA3.1/V5/His.

In a fourth aspect, the invention provides a cell which comprises a vector according to the third aspect of the invention.

In one embodiment, the cell of the fourth aspect of the invention is an E.coli cell. Preferably, the E. coli is BM25.8. In another embodiment, the cell is a COS-7 cell.

In a fifth aspect, the invention provides a method for making a peptide according to the first aspect of the invention which comprises the step of maintaining a cell according to the fourth aspect of the invention in conditions in which the peptide is expressed by the cell.

In one embodiment, the method of the fifth aspect of the invention comprises the further step of isolating the peptide.

In a sixth aspect, the invention provides a peptide when produced by the method of the fifth aspect of the invention.

In a seventh aspect, the invention provides a composition comprising a peptide according to the first or sixth aspect of the invention and a pharmaceutically acceptable carrier.

In an eighth aspect, the invention provides an antibody which is capable of binding a peptide according to the first or sixth aspect of the invention.

In one embodiment, the antibody of the eighth aspect of the invention is secreted by a hybridoma cell.

In a ninth aspect, the invention provides a hybridoma cell which secretes an antibody according the eighth aspect of the invention.

In a tenth aspect, the invention provides a method of cleaving a molecule which comprises a di-peptide sequence Ala-Pro at a peptide bond which is C-terminal adjacent to proline in the di-peptide, the method comprising maintaining the molecule in the presence of a peptide according to the first aspect or the sixth aspect of the invention so that the peptide bond C-terminal adjacent to proline in the di-peptide is cleaved.

In one embodiment of the tenth aspect of the invention, the molecule further comprises the di-peptide sequence, Gly-Pro.

In an eleventh aspect the invention provides a kit comprising the peptide of the first or sixth aspects of the invention, or an antibody according to the eighth aspect of the invention.

BRIEF DESCRIPTION OF THE FIGURES

Fig 1. Cloning strategy for isolating full-length DPP4L1 cDNA and the alternative splicing variants of DPP4L1 observed. Representation of three splice variants is shown including loss of serine recognition site by one splice variant (T8).

Fig 2. Nucleotide sequence and amino acid sequence of human DPP4L1. The nucleotide and predicted one letter code amino acid sequence are shown. This sequence shows no putative membrane spanning domain (deduced from hydrophobicity plots) or potential N-linked glycosylation sites. The putative serine recognition site and aspartic acid and histidine which form the SER-ASP-HIS catalytic domain are shaded. Base pairs are numbered in the right margin.

Fig 3. Alignment of the predicted protein sequence of DPP4L1 with human DPP4 and *C elegans* homologue. The amino acid sequences were aligned using PileUp alignment program in GCG. Amino acid residues identical in all three proteins are boxed.

Fig 4. Northern Blot analysis of DPP4L1 expression. Human multiple tissue Northern blots (CLONTECH) containing 2 ug per lane of poly A RNA were hybridized with a ^{32}P labeled DPP4L1 probe at $68^{0}C$ and washed at high stringency. The autoradiograph was exposed for 1 day at $-70^{\circ}C$ with a BIOMAX MS screen. Molecular mass markers are indicated in base pairs on the left side of each autoradiogram.

Fig 4a. Master RNA (CLONTECH) blot of poly A RNA was hybridized with a ^{32}P labeled DPP4L1 probe at 65°C and washed at high stringency. The autoradiograph was exposed for 3 day at -70°C with BIOMAX MS screen. DPP4L1 mRNA was detected in all tissues examined.

Fig 5. Chromosomal localization of human DPP4L1. Metaphase showing FISH with the biotinylated DPP4L1 cDNA probe. Normal male chromosomes stained with DAPI. Hybridization sites on chromosome 15 are indicated by an arrow.

Fig 6. Western blot analysis of transfected cell lines. Analysis of lysates of stable cell lines. DPP4L1 protein was seen in DPP4L1 /V5/His stable cell line but not in DPP4 or vector only stable cell lines. The electrophoretic mobility of the protein was not altered when samples were boiled. The band of greater mobility was probably a breakdown product of intact DPP4L1.

Fig 7. Human DPP4L1 confered Ala-Pro DPP activity upon COS cells transfected with DPP4L1 cDNA.

Fig. 8 Detection of DDP4L1 expression in COS-7 cells by fluorescent staining and phase contrast microscopy.

DETAILED DESCRIPTION OF THE INVENTION

EXAMPLES

General

Restriction enzymes and other enzymes used in cloning were obtained from Boehringer Mannheim Roche. Standard molecular biology techniques were used (30) unless indicated otherwise.

An EST clone (GENBANK™ accession number AA417787) was obtained from American Type Culture Collection. The DNA insert of this clone was sequenced on both strands using

automated sequencing at SUPAMAC (Sydney, Australia).

DPP4L1 Cloning

ESTAA417787 was used to design forward (caa ata gaa att gac gat cag gtg) and reverse (tct tga agg tagtgc aaa aga tgc) DPP4L1 primers for polymerase chain reaction (PCR) from ESTAA417787. The PCR conditions were as follows: 94°C for 5 min, followed by 35 cycles of 94°C for 1 minute, 55°C for 30 sec and 70°C for 1 min. This 484 bp PCR product was gel purified, 32 P- labeled using Megaprime Labeling Kit (Amersham Pharmacia Biotec, UK) and hybridized to a Master RNA blot (CLONTECH, Palo Alto, CA, USA) that contained poly A^+ from 50 adult and fetal tissues immobilized in dots as per manufacturers' instructions. This Master RNA blot was also probed with DPP4 for comparison of mRNA tissue expression.

The forward and reverse DPP4L1 primers were used for PCR to screen a human placental λ STRETCH PLUS library (CLONTECH, Palo Alto, CA, USA) for the presence of DPP4L1 cDNA in the library. The library was then screened by standard molecular biology techniques. After primary screening, 23 clones were selected for secondary screening, after which 22 remained positive. For the tertiary screen the clones contained in λTripleEx were converted into pTriplEx plasmids and transformed into BM25.8 *E. coli* recipient bacteria. The plated bacteria were screened and it was confirmed that all 22 clones were positive. Two of these clones, T8 and T21 were selected for further study.

5'RACE (Rapid amplification of cDNA ends)

A 5' RACE Version 2.0 kit (Gibco BRL, Life technologies) was applied on activated T cell (ATC) and placental RNA as prescribed in the kit instructions. The T8 DNA sequence was used to design GSP 1 (TCC TTC CTT CAG CAT CAA TC) and GSP2 (CTT AAA AGT GAC TTT AGG ATT TGC TGT ACC). 5' RACE PCR

products were cloned into pGEM-T Easy®Vector (Promega Co., Madison, WI, USA) and sequenced by primer walking.

Confirmation of identity of RACE product

Reverse transcriptase PCR was carried out on ATC RNA using DPP4L1-pr23 (GGA AGA AGA TGC CAG ATC AGC TGG) and DPP4L1-pr19r (TCC GTG TAT CCT GTA TCA TAG AAG) to span across the junction between the RACE product and the EST and library clones. Two gel purified products ATCd3-2-1 (1603bp) and ATC3-3-10 (1077bp) were cloned into pGEM-T Easy® (Promega Co., Madison, WI, USA) and sequenced.

Subcloning of DPP4L1 cDNA into a pcDNA3.1/V5/His Expression Vector

The ATC RACE product, the ATCd3-2-1 (1603bp) junction fragment and the library clone T21 were joined together and cloned into the expression vector pcDNA3.1/V5/His A (Invitrogen) to form a DPP4L1 cDNA of 3.1 kb with an open reading frame of 882 aa. The first construct was made using three sequential cloning steps. Firstly, a *Eco RV/Xba I* fragment of T21 (containing 3' DPP4L1, stop codon and 3' untranslated region on DPP4L1 cDNA) was ligated into the vector pcDNA3.1/V5/His A which had been digested with *Eco RV/Xba I*. An *Eco RI/Eco RV* fragment of ATCd3-3-1 was then added to this construct digested with *Eco RI/Eco RV*. Finally the RACE product was cut with *Eco RI* and cloned into the *Eco RI* site of the previous construct to form the complete 3.1 kb DPP4L1 cDNA. This construct pcDNA3.1-DPP4L1 expressed protein with no detectable tag. In addition the stop codon in the DPP4L1 expression construct in pcDNA3.1/V5/His V5 was genetically altered using PCR to create a C-terminal fusion with the V5 and His tag contained in the vector. This construct was named pcDNA3.1-DPP4L1/V5/His. All expression constructs subcloned into pcDNA3.1/V5/His were verified by full sequence analysis.

DPP4L1 gene expression by Northern Blot

Human multiple tissue Northern blots (CLONTECH) containing 2 ug of poly A^+ RNA were prehybridized in Express Hybridization solution (CLONTECH) for 30 min at 68°C. Both the DPP4L1 484 bp product and the 5' RACE ATC product were radiolabeled using a Megaprime Labeling kit (Amersham Pharmacia Biotech) and [32P]dCTP (NEN Dupont). Unincorporated label was removed using a NICK column (Amersham Pharmacia Biotech) and the denatured probe was incubated for 2 hrs at 68°C in Express Hybridization solution. Washes were performed at high stringency and blots exposed to BIOMAX MS film for overnight with a BIOMX MS screen at -70°C.

DPP4L1 expression by RT-PCR

Reverse transcriptase PCR was performed on human ATC RNA, human placental RNA and human liver RNA using TED primers DPP4L1/pr3 (GCA CTA CCT TCA AGA AAA CCT TGG) and DPP4L1/pr20R (TAT GGT ATT GCT GGG TCT CTC AGG) to give a 293 bp product.

Transient Transfection into COS cells

Monkey kidney fibroblast (COS-7) cells (ATCC, CRL-1651), were grown in Dulbecco's MEM medium supplemented with 10% fetal calf serum and 2mM glutamine. A subconfluent 75 cm^2 flask of COS cells was transfected using 15 ug DNA and 48 ul Fugene-6 (Roche, Palo Alto, CA, USA) following the manufacturer's instructions. Cells were incubated for 72 hrs before harvesting. For making stable cell lines, Geneticin (G418, Gibco BRL) was added 24 hrs after transfection and cells were maintained and grown continuously in media containing G418 selection.

Determination of DPPactivity of DPP4L1

DPP4 enzyme assays were performed on trypsin/EDTA-harvested

COS-7 cells 72 hrs after transfection and used Gly-Pro-p-nitroanilide-p-toluene sulfonate salt (Sigma, St Louis, MO, USA) [Duke-Cohan, 1996 #1406], Ala-Pro-p-nitroanilide HCl (Bachem, Switzerland) or Gly-Arg- p-nitroanilide HCl (Sigma) as the substrates. Transfected cells were lyzed by sonication then incubated at 20,000 cell equivalents per well in 70µl phosphate buffer, pH7.0 for 40 minutes at 37^{0}C. Absorbances at 690nm were subtracted from absorbances at 405nm to increase the specificity of measurements. Analyses of Michaelis Menten kinetics used KaleidaGraph (Hearne Scientific Software). Assays were performed in triplicate on two transfections.

Chromosomal localization of DPP4L1 by Fluorescence in situ Hybridization (FISH) analysis

DPP4L1 was localized using two different probes, the DPP4L1 EST and the T8 clone. The probes were nick-translated with biotin-14-dATP and hybridized *in situ* at a final concentration of 10ng/ul to metaphases from two normal males. The FISH method was modified from that previously described (31) in that chromosomes were stained before analysis with both propidium iodide (as counterstain) and DAPI (for chromosomal identification). Images of metaphase preparations were captured by a cooled CCD camera using the Cyto Vision Ultra image collection and enhancement system (Applied Imaging Int Ltd). FISH signals and the DAPI banding pattern were merged for figure preparation.

Molecular cloning and sequence analysis of DPP4L1

The insert in ATCC EST AA417787 was 805 bp in length, containing 537 bp of coding sequence, a TAA stop codon and 267 bp of 3' noncoding sequence (Figure 1).

The hybridization of the Master RNA blot revealed that the gene comprising ESTAA417787 has ubiquitous tissue expression, with high levels of expression in testis and

placenta. Based on this expression pattern, a placental cDNA library was screened with a 484 bp PCR product produced by the forward and reverse DPP4L1 primers. Sequence homology analysis revealed that only 2 of 23 clones contained 5' sequence additional to the sequence of ESTAA417787. These cDNA clones were designated T8 and T21, and were 1.7 kb and 1.2 kb respectively (Figure 1). In addition, comparison of these sequences to ESTAA417787 revealed that T8 cDNA lacked a 153 bp (51aa) region that was present in T21 cDNA and ESTAA417787. This deletion would result in the loss of the catalytic serine (GWSYGG) in T8 cDNA. Many of the other clones characterized appeared to contain unrelated sequence which are probably intronic sequences as a result of incomplete splicing.

The 5' RACE technique was utilized on both ATC RNA and placental RNA to obtain the 5' of end of the DPP4L1 gene. The RACE product obtained from activated T cell RNA was 0.2 kb larger than that from placental RNA but otherwise identical (Figure 1). The first methionine within a Kozak sequence was found 211 bp from the 5' end of the activated T cell RACE product. This 5' 211bp region was 70.5 % GC rich and contained a number of potential promoter and enhancer elements (Sp1, Ap1 and ETF sites) and so was deduced to be the 5' flanking region of the DPP4L1 gene. In order to confirm the identity of the 5' RACE product as the 5' end of DPP4L1, RT-PCR was carried out to span across the junction between the RACE product and T8 cDNA library clone. The RT-PCR on ATC RNA produced two clones ATCd3-2-1 and ATC3-3-10 (Figure 1). Compared to T8 and T21, both clones had an additional insert region of 144bp (48 aa) immediately adjacent to the splice site of T8. Sequence homology analysis of this additional insert region found a homologous region in both the *C. elegans* homologue and DPP4. This clearly showed that T8 and T21 library clones represented splice variants of DPP4L1. The smaller clone

ATCd3-3-10 was also found to represent another splice variant of DPP4L1 as it contained a 516 bp deletion at the 5' end which would result in a deletion of 175 aa. At this point it is unclear about the biological significance of three different splice variants observed.

A full-length DPP4L1 clone was created using the larger RACE product, ATC3-2-1 and the T21 library clone. This generated a putative DPP4L1 cDNA of 3.1 kb (including 5' and 3' untranslated regions) with an open reading frame of 882 aa for further sequence analysis and examining DPP4L1 function. This 882 putative DPP4L1 protein contained no N-linked glycosylation sites and Kyte-Doolittle hydrophobicity analyses revealed it lacked a transmembrane domain, unlike DPP4, FAP and DPP6. Thus it is likely that DPP4L1 is a cytoplasmic protein (Figure 2). The predicted DPP4L1 protein shared 51 % amino acid similarity and 27 % amino acid identity with human DPP4; the C termini of these proteins exhibited the most homology (Figure 3).

Tissue distribution of DPP4L1 as determined by Master RNA and Northern Blot

A master RNA blot was probed with a 484 nt PCR product produced by the forward and reverse DDP4L1 pimers as mentioned previously. The mRNA tissue expression of DPP4L1 was ubiquitous in all human adult and fetal tissues. A similar ubiquitous expression pattern was observed using DPP4 cDNA as a probe (data not shown). However, by visual assessment the greatest levels of expression using each gene specific probe were in different tissues. The most intense signals using the DPP4L1 probe were in testis followed by placenta whereas the most intense signals using the DPP4 probe were in salivary gland and prostate gland followed by placenta (data not shown). The probes did not bind any of the negative controls on the blot.

Northern blot analysis was performed on mRNA derived from different human tissues (Figure 4). Two DPP4L1 specific probes indicated the presence of transcripts in all tissues examined. A transcript approximately 3.0 kb in size consistent with the approximate expected size of DPP4L1 message was detected only in the testis. However, two transcripts of 8.0 and 5.0 kb respectively were present in testis, spleen, peripheral blood leukocytes and ovary at high levels; in prostrate, small intestine, and colonic mucosa at moderate levels; and in the thymus at lower levels. The Multiple tissue Northern blot was also probed with radiolabeled human β-actin probe and a common 2.0 kb transcript was seen in all tissues (Figure 4).

Expression and functional activity of DPP4L1

To assess the function of DPP4L1 protein, the full length DPP4L1 cDNA of 3.1 kb was cloned into the *Xba I* site of pcDNA3.1A/V5/His expression vector to produce two constructs. The first construct, pcDNA3.1-DPP4L1, expressed DPP4L1 protein on its own whilst the second construct, pcDNA3.1-DPP4L1/V5/His expressed a protein with the V5 epitope and His tag fused to the C-terminus of DPP4L1 to facilitate analysis of protein expression. Mammalian expression constructs were stably transfected into COS-7 cells and cellular sonicates prepared. Consistent with the molecular weight predicted from the amino acid sequence a 100 kDa monomer was detected by Western blotting of stable DPP4L1/V5/His expressing cells (Figure 6). DPP4L1/V5/His protein was detected in the cytoplasmic compartment but not on the surface of ethanol fixed stable DPP4L1/V5/His expressing COS cells, using the anti-V5 mAb. Due to homology between DPP4 and DPP4L1 cell lysates were examined for serine protease activity. Expression of DPP4 with and without the V5 and His tags in COS cells was performed as a positive control and to establish the working conditions of the assay. Homogenates of vector-only transfections were

used in parallel as negative controls. Extracts of DPP4L1-transfected cells hydrolyzed Ala-Pro-p-nitroanilide but not Gly-Pro-p-nitroanlide, Gly-Arg-p-nitroanilide, Gly-Pro-toluene sulphonate or Gly-Pro-7-amino-4-trifluoromethyl coumarin.

Chromosomal localization of DPP4L1

Two probes were used for FISH analysis, ESTAA417787 and the T8 clone from the placental library. Seventeen metaphases from the first normal male were examined for fluorescent signal. All of these metaphases showed signal on one or both chromatids of 15 at band q22 (Figure 5). There were a total of 2 non-specific background dots observed in these metaphases. A similar result was obtained form the hybridization of the probe to 15 metaphases from the second normal male (data not shown).

We describe a novel human POP that we have called DPP4L1 protein and the gene encoding it. Analysis of the open reading frame of the complete DPP4L1 cDNA sequence suggested that it is a cytoplasmic protein. Hydropathy analysis indicated that in contrast to DPP4, FAP and DDP6 genes, DPP4L1 does not contain a short hydrophobic region to act as membrane spanning domain. Human DPP4, FAP and DPP6 contain between 6 and 9 potential N-glycosylation sites in their amino acid sequence. A similar examination of DPP4L1 cDNA sequence revealed that it had no sites of this type which was further indication that DPP4L1 was a cytoplasmic protein. The detection of tagged DPP4L1 protein in the cytoplasmic compartment but not on the surface of transiently transfection COS cells, using the anti-V5 mAb, further suggested that DPP4L1 is a cytoplasmic protein.

The most significant homology between DPP4L1 and DPP4 is in the C termini where the three catalytic residues Ser, Asp and His are located. By homology with DPP 4, DPP4L1 is a

member of the DPP 4-like gene family, a member of the POP family and a member of the α/β hydrolase fold family(32). The catalytic residues in DPP4L1 that potentially form the charge-relay system are Ser^{739}, Asp^{817} and His^{849}. Transfection experiments were performed with constructs of DPP4L1 cDNA to demonstrate its ability to behave as a serine protease and exhibit DPP enzyme activity. DPP4L1 cDNA constructs conferred DPP enzyme activity to cellular homogenates as demonstrated by their ability to hydrolyze the substrate Ala-Pro. However, constructs of DPP4L1 did not confer activity against Gly-Pro upon transfected COS cells, indicating that DPP4L1 has a different substrate specificity to DPP4. The physiological role of hydrolysis of Ala-Pro is unknown.

When DPP4 is expressed on the surface of T cells it is know as the cell surface antigen CD26. CD26-negative cell lines have been shown to have residual DPP4 activity, indicating the existence of alternative peptidase with DPP4 activity. DPP4β is protein which shows a peptidase activity similar to DPP4 and has been purified from the CD26-negative cell line C8166 (27, 28). Purified DPP4β, cleaves Gly-Pro substrate and is a glycosylated protein that exists on the cell surface as 70-80 kDa monomer. Therefore, according to the substrate specificity, cellular localization and biochemical properties DPP4L1 is novel DPP distinct from DPP4β

During the cloning of DPP4L1 it became apparent that at least three alternately spliced transcripts of DPP4L1 other than full-length are present in tissues examined. The biological significance of such transcripts is so far unknown. None of the three splice variants result in a frame shift or premature protein termination so can potentially produce intact but truncated DPP4L1 proteins. Two of the three splice variants contain all the catalytic

triad residues and thus may still produce proteins with DPP activity. It is possible that expression of these sequences may be used to regulate the levels of active protein. In addition, analysis of DPP4L1 tissue distribution by Northern hybridization revealed a number of differently sized transcripts. However the size of these transcripts did not concur with those expected to be seen from alternate splicing. The predicted size of alternate spliced variants of DPP4L1 would range in size from 2.6 -3.1 kb whereas the large transcripts seen in most tissues examined in the Northern blot were 8.5 and 5.0 kb in size respectively. These transcript sizes are much larger than the 3.1 kb transcript predicted for DPP4L1 from the cloning strategy. These large transcripts may contain 5' and 3' untranslated sequences and therefore may still encode functional DPP4L1 protein. However, it is also possible that these transcripts represent incompletely spliced mRNA transcripts and therefore do not produce intact DPP4L1 protein. Further work will determine the role of DPP4L1 in different tissues and whether alternative splicing has any biological role.

Using FISH analysis to determine the chromosomal localization of DPP4L1, we observed a signal on chromosome 15q22 for DPP4L1. Both DPP4 and FAP have been localized to the long arm of chromosome 2, 2q24.3 (33) and 2q23 (34) respectively. DPP6 which is further in sequence from DPP 4 and FAP was localized to chromosome 7(21). The localization of DPP4L1 to 15q22 predicts that DPP4-like gene family members could be spread throughout the human genome, and may be present on other chromosomes. The structure of a gene in *C. elegans* which encodes an amino acid sequence which is homologous to DDP4L1 has 19 exons spanning 5.3 kb. In *C. elegans* DPP4L1, the serine recognition site, GWSWGG, is found in exon 16 and does not span two exons as found in the genes for *C. elegans* and human DPP 4(6), and human and

mouse FAP (25). The serine recognition site for *C. elegans* PEP is also found in one exon therefore this arrangement may be representative of the ancestral POP gene and the arrangement in DPP 4 and FAP may have resulted from divergent evolution from this ancestral gene.

In summary we have identified and characterized a novel human POP DPP4L1 that exhibits DPP activity and the gene encoding it.

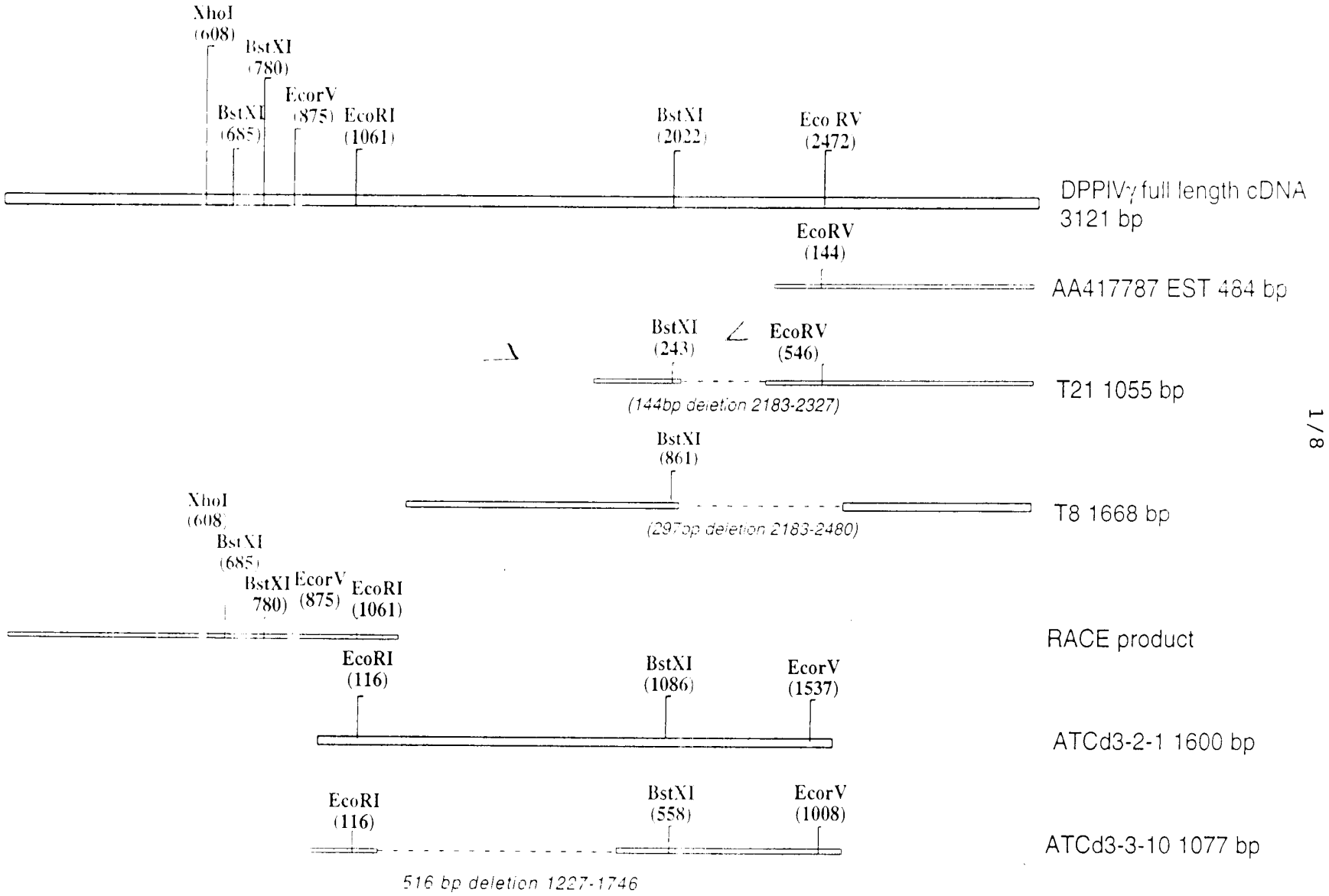
XhoI (608)
BstXI (780)
EcorV (875)
BstXI (685)
EcoRI (1061)
BstXI (2022)
Eco RV (2472)
DPPIVγ full length cDNA 3121 bp
EcoRV (144)
AA417787 EST 484 bp
BstXI (243)
EcoRV (546)
T21 1055 bp
(144bp deletion 2183-2327)
BstXI (861)
T8 1668 bp
(297bp deletion 2183-2480)
XhoI (608)
BstXI (685)
BstXI 780)
EcorV (875)
EcoRI (1061)
RACE product
EcoRI (116)
BstXI (1086)
EcorV (1537)
ATCd3-2-1 1600 bp
EcoRI (116)
BstXI (558)
EcorV (1008)
ATCd3-3-10 1077 bp
516 bp deletion 1227-1746

FIGURE 1

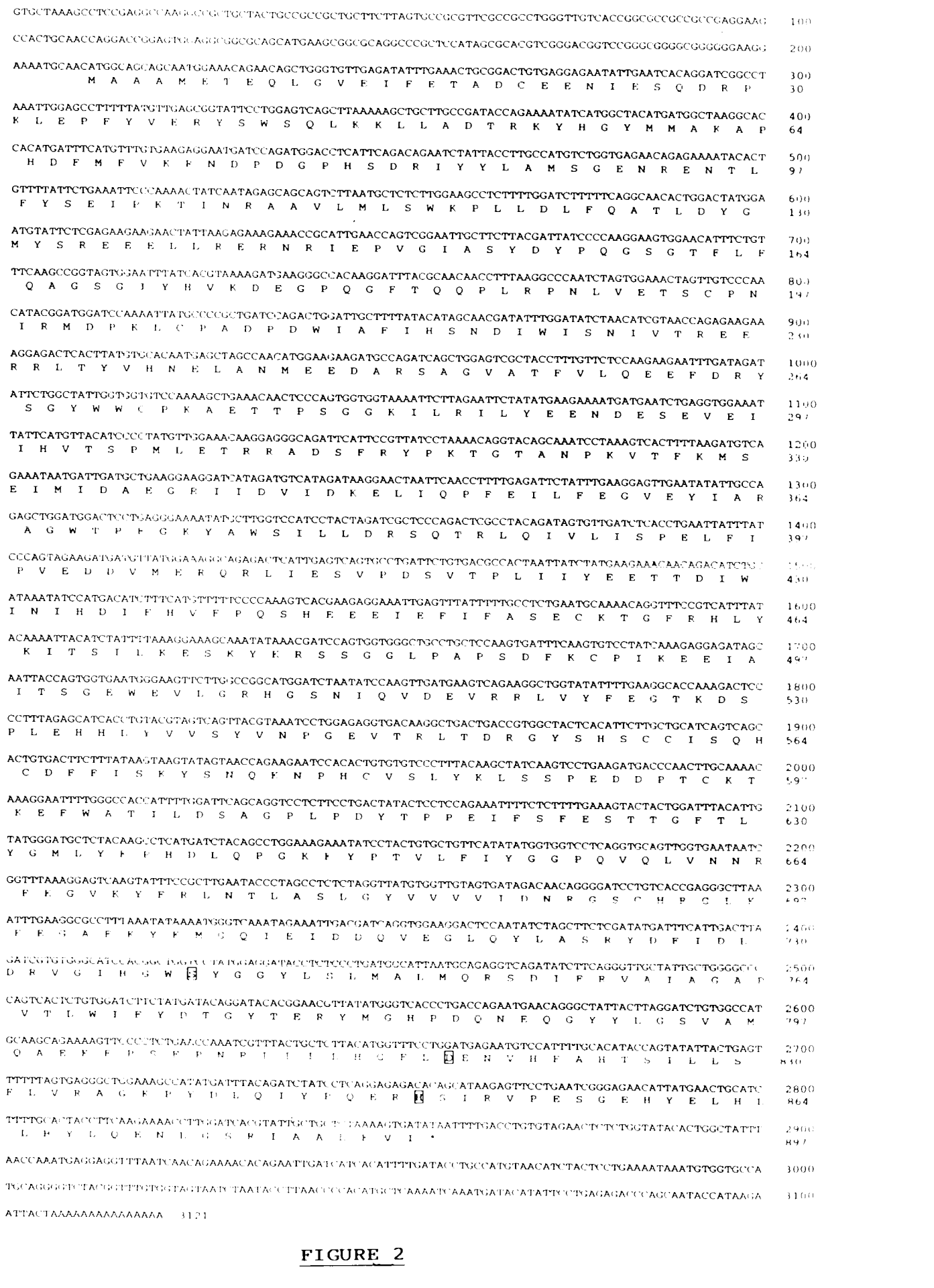

FIGURE 2

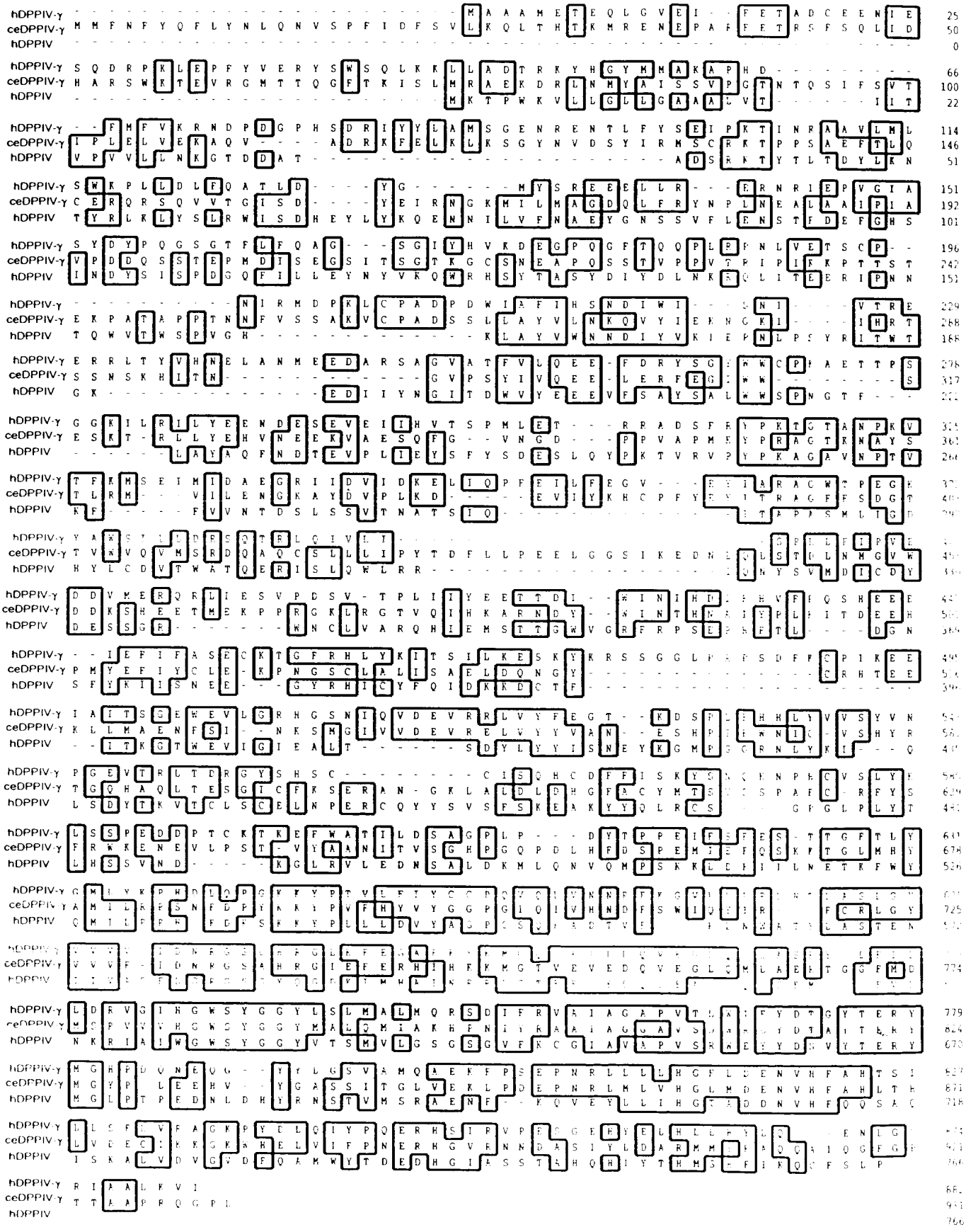

FIGURE 3

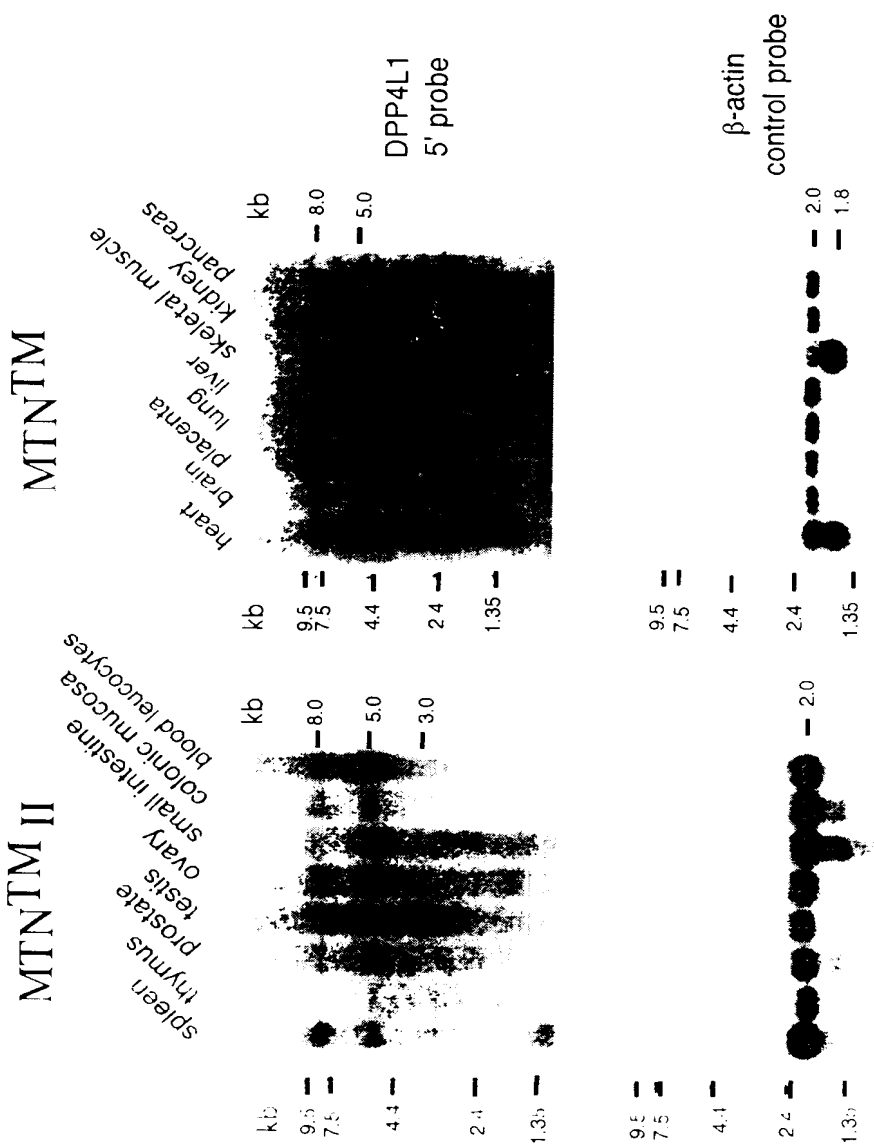
MTN TM II
MTN TM
spleen
thymus
prostate
testis
ovary
small intestine
colonic mucosa
blood leucocytes
heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas
kb
8.0
5.0
3.0
2.0
9.5
7.5
4.4
2.4
1.35
DPP4L1
5' probe
β-actin
control probe
2.0
1.8

FIGURE 4

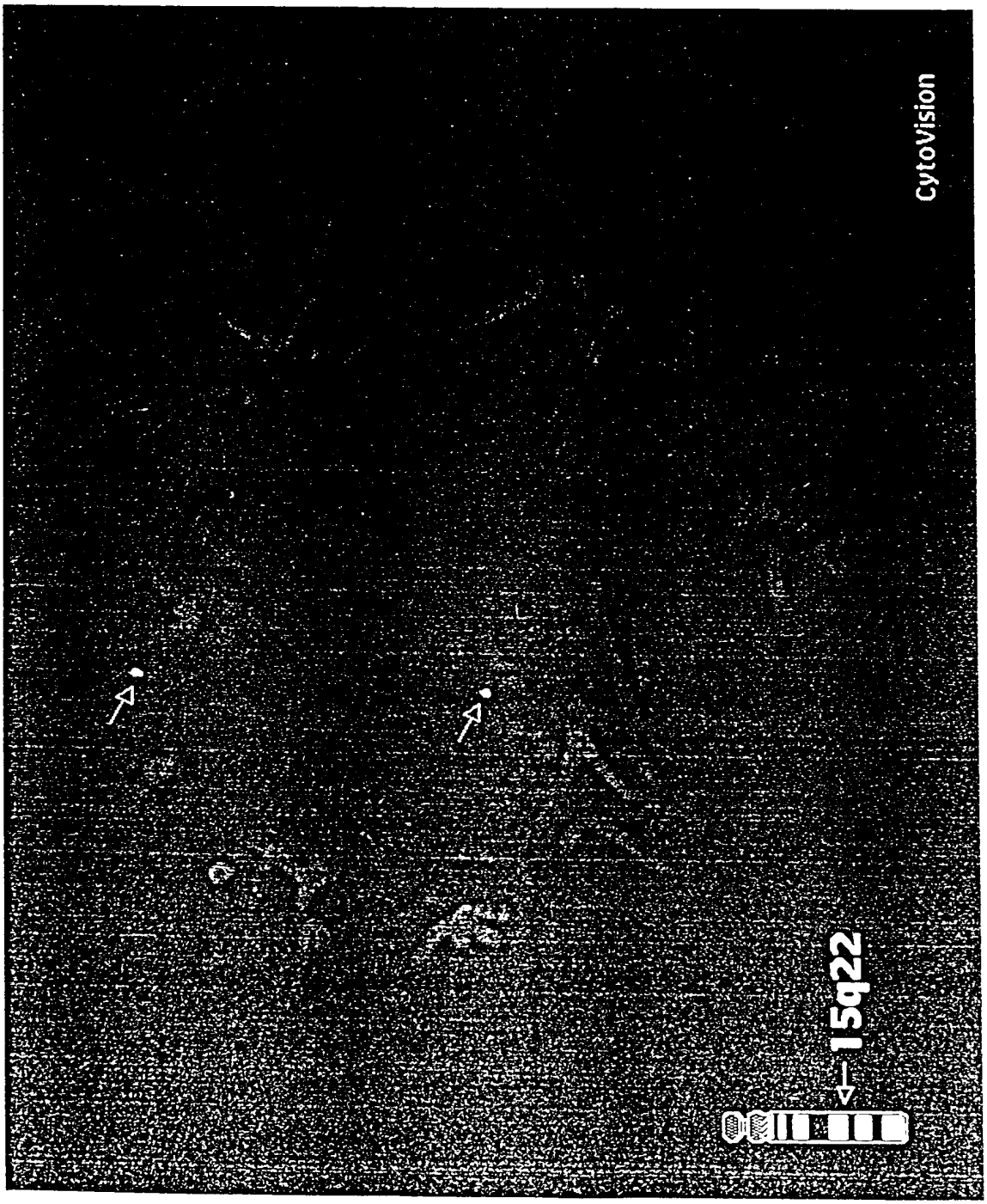

FIGURE 5

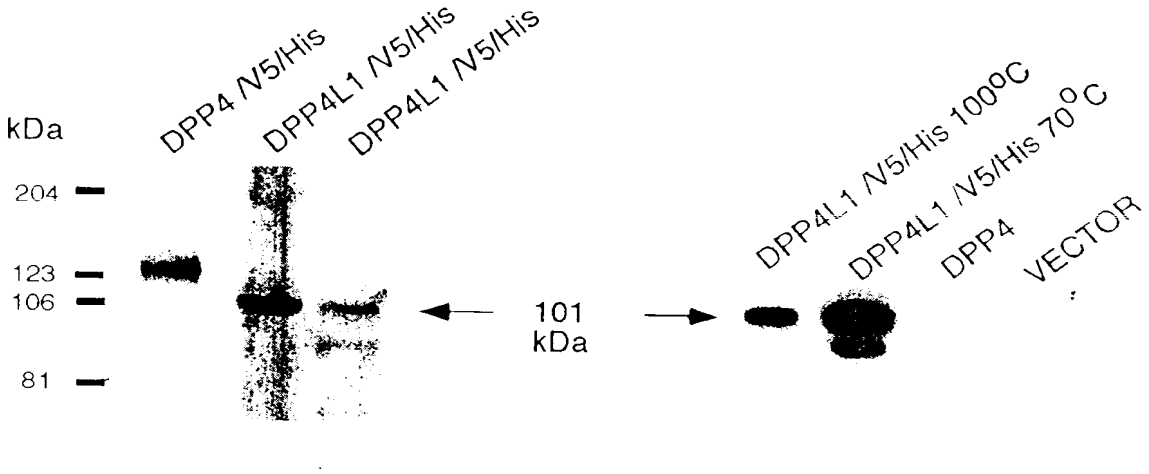
kDa
DPP4 /V5/His
DPP4L1 /V5/His
DPP4L1 /V5/His
204
123
106
81
101
kDa
DPP4L1 /V5/His 100°C
DPP4L1 /V5/His 70°C
DPP4
VECTOR

FIGURE 6

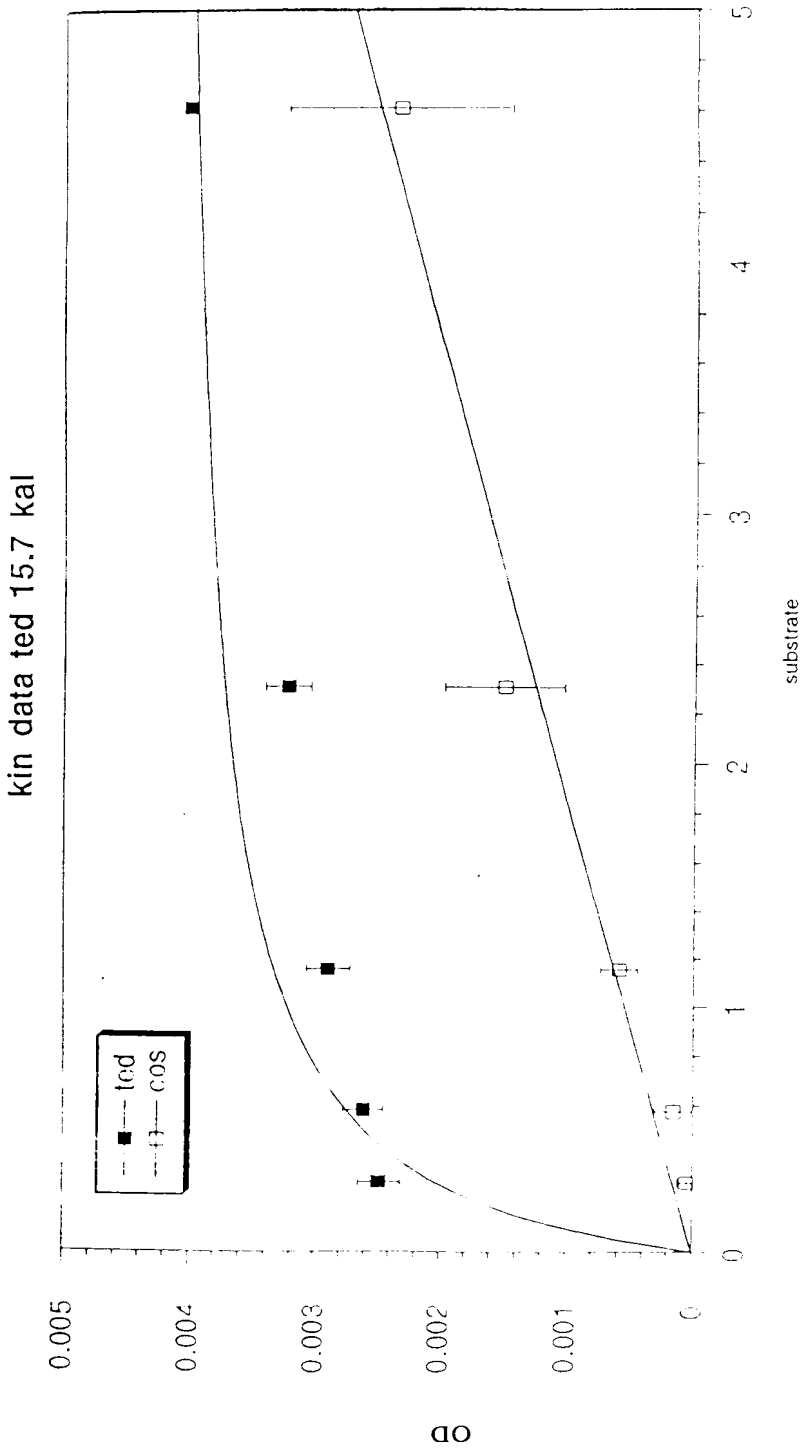
kin data ted 15.7 kal
ted
cos
OD
0.005
0.004
0.003
0.002
0.001
0
1
2
3
4
5
substrate

FIGURE 7

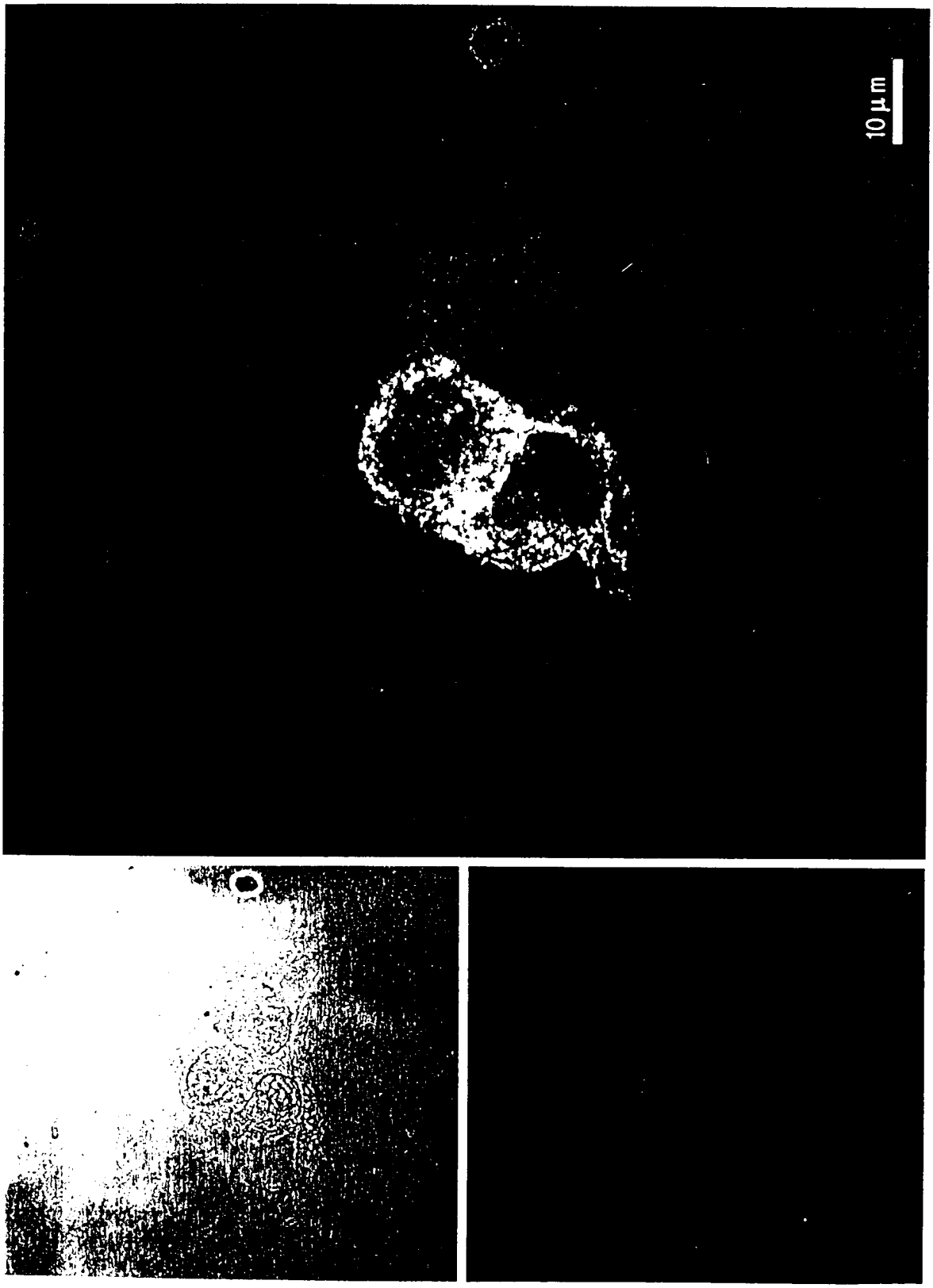

FIGURE 8